International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-12, Issue-3; Mar, 2025
Journal Home Page Available: ***https://ijaers.com/***
Article DOI: *https://dx.doi.org/10.22161/ijaers.123.1*

Understanding Legislative Activities in Brazil: Analysis of Open Data from the Chamber of Deputies

Compreensão das Atividades Legislativas no Brasil: Análise dos Dados Abertos da Câmara dos Deputados

Milena Carneiro Rios de Oliveira[1], Anna Paula de Sousa Parente Rodrigues[2]

[1]Department of Computer Science, Federal University of Tocantins, Brazil
anna.rodrigues@uft.edu.br
[2]Department of Computer Science, Federal University of Tocantins, Brazil
milena.rios@uft.edu.br

Received: 27 Jan 2024,

Receive in revised form: 25 Feb 2025,

Accepted: 02 Mar 2025,

Available online: 07 Mar 2025

Keywords**—Chamber of Deputies, CRISP-DM, Data Mining, K-Means.**

***Palavras-chave* — Câmara dos Deputados, CRISP-DM, Mineração de Dados, K-Means.**

Abstract**—** *The growing evolution of technological means as gears of modern life become responsible for the constant generation of data, which in turn imply the most diverse aspects of society. In this sense, data analysis is a powerful device when it comes to predicting and describing information, in order to improve decision-making. Thus, the insertion of open data in the space of public management enables greater transparency with the citizen, making social control and the consolidation of democracy possible. Based on this bias, the present study aims to implement Data Mining techniques on the open data of the Chamber of Deputies, aiming to broaden the understanding of legislative activities and present the knowledge discovered from the relationships and patterns found in the datasets. Therefore, this is an exploratory study, based on the CRISP-DM methodology, covering the steps of understanding the business, understanding and preparing data, modeling, evaluation and implementation, corresponding to post-processing.*

Resumo— *A crescente evolução dos meios tecnológicos como engrenagens da vida moderna tornam-se responsáveis pela constante geração de dados, que por sua vez implicam os mais diversos aspectos da sociedade. Nesse sentido, a análise de dados é um poderoso dispositivo quando se trata de prever e descrever informações, a fim de aprimorar a tomada de decisão. Assim, a inserção de dados abertos no espaço da gestão pública possibilita maior transparência com o cidadão, tornando possível o controle social e a consolidação da democracia. Com base nesse viés, o presente estudo tem como objetivo implementar técnicas de Mineração de Dados sobre os dados abertos da Câmara dos Deputados, visando ampliar a compreensão das atividades legislativas e apresentar o conhecimento descoberto a partir das relações e padrões encontrados nos conjuntos de dados. Portanto, trata-se de um estudo exploratório, baseado na metodologia CRISP-DM, abrangendo as etapas de compreensão do negócio, compreensão e preparação dos*

dados, modelagem, avaliação e implementação, correspondendo ao pós-processamento.

I. INTRODUÇÃO

A evolução tecnológica tem desempenhado um papel crucial na transformação da sociedade moderna, impulsionando a geração contínua de dados que afetam diversos aspectos da vida cotidiana. Neste contexto, a análise de dados emerge como uma ferramenta essencial para a previsão e descrição de informações, aprimorando a tomada de decisões em múltiplos setores [1].

A transparência governamental, facilitada pela disponibilização de dados abertos, é um dos campos que mais se beneficia dessa evolução, promovendo o controle social e fortalecendo a democracia [2].

Este estudo foca na aplicação de técnicas de mineração de dados sobre os dados abertos da Câmara dos Deputados do Brasil. O objetivo central é ampliar a compreensão das atividades legislativas, através da aplicação do algoritmo k-means, revelando padrões e relações que possam informar melhor o público e os tomadores de decisão. A pesquisa adota a metodologia CRISP-DM, um modelo robusto para a análise de dados, guiando o processo desde a compreensão do negócio até a implementação dos resultados [3].

Este estudo parte da hipótese de que os dados abertos da Câmara dos Deputados possibilitam um acompanhamento transparente do exercício parlamentar dos governantes eleitos, sendo este um direito garantido constitucionalmente e que proporciona subsídios ao cidadão a fim de ampliar a participação social com o acesso direto e o reconhecimento da honestidade e integridade, visando alcançar maior visibilidade sobre ações e decisões governamentais. Contudo, ainda que com o aumento da transparência legislativa e a disponibilização de dados em tempo real pela Câmara dos Deputados, há uma carência de maior ênfase na obtenção de informações a partir de análises mais significativas dos dados expostos e melhor apresentação dos resultados, o que interfere diretamente na capacidade de fiscalização, monitoramento e discussão na sociedade [4] e [5].

Ao investigar o uso de dados abertos no contexto da gestão pública, este artigo não apenas contribui para o entendimento das práticas legislativas, mas também destaca a importância da ciência de dados como um catalisador para a transparência e a participação cidadã. A análise dos dados abertos da Câmara dos Deputados serve como um estudo de caso para demonstrar o potencial transformador da tecnologia na governança pública.

II. TRABALHOS RELACIONADOS

Em [3] é investigada as dimensões dos mecanismos digitais de interação e participação política oferecidos pelos parlamentos à sociedade. O estudo foca na compreensão das iniciativas digitais de transparência, com um olhar específico sobre a Câmara dos Deputados do Brasil. A autora analisa como essas ferramentas digitais podem facilitar a participação cidadã e a transparência governamental, destacando a importância de um acesso eficiente e compreensível às informações legislativas. O trabalho também enfatiza a E-Democracia advinda da ascensão de Tecnologias da Informação e Comunicação (TICs), convergindo para a participação política através da relação existente entre política e internet e possibilitada por meio dos portais legislativos e os recursos oferecidos. Nesse viés, compreende-se a história e a motivação dos portais legislativos, além das oportunidades cedidas aos usuários.

Já [6] apresenta como objetivos, desenvolver um diagnóstico sobre a demanda por dados abertos da Câmara dos Deputados e verificar a sua contribuição para a transparência legislativa. A metodologia realizada buscou, inicialmente, analisar através do Google Analytics as cidades brasileiras que mais acessaram o portal de transparência da câmara no ano de 2015, além da quantidade de visualizações em cada serviço, seja na solicitação dos deputados em exercício, das características dos deputados, dos partidos, etc., indicando a porcentagem em relação ao todo. Assim, o intuito dos autores foi demonstrar resultados de diferentes análises quanto ao uso do portal, permitindo verificar as principais em relação à busca de informações pelos usuários, melhorando o desenvolvimento das aplicações sobre os dados disponibilizados.

Por fim, [7] detalha a aplicação da metodologia CRISP-DM (Cross-Industry Standard Process for Data Mining) na análise de dados legislativos. A pesquisa demonstra como cada etapa do CRISP-DM, desde a compreensão do negócio até a implantação, pode ser aplicada para analisar dados abertos da Câmara dos Deputados. O estudo enfatiza a importância de uma abordagem estruturada para a análise de dados, permitindo a descoberta de padrões e relações significativas nos dados legislativos. Além disso, a pesquisa ilustra a utilização de algoritmos de mineração de dados, como o k-means, para modelar e interpretar os dados, oferecendo uma visão aprofundada das atividades legislativas e facilitando a tomada de decisões informadas.

Os trabalhos relacionados contribuíram para a pesquisa atual, oferecem uma base teórica sólida sobre mecanismos digitais de interação e participação política, além de servirem como guias metodológicos ajudando a refinar as etapas de modelagem e análise de dados no trabalho atual.

III. METODOLOGIA

As etapas de desenvolvimento deste trabalho seguiram o padrão do modelo de processo Crisp-DM. Sua escolha se deu uma vez que ela oferece uma rota completa de etapas a fim de obter melhor entendimento e condução de projetos, sendo idealizado em 1996 por quatro líderes de empresas pioneiras na área: Daimler-Benz, Integral Solutions Ltd, NCR e OHRA [8]. Tal metodologia é composta pelas seguintes etapas:

Compreensão do negócio

Nessa etapa, buscou-se conhecer o domínio da aplicação, sendo este o conjunto de dados públicos sobre a Câmara dos Deputados, incluindo informações sobre legislaturas, deputados, partidos, blocos, órgãos, frentes parlamentares, eventos, proposições, votações, referências e atualizações, totalizando 28 conjuntos distintos de dados. Além disso, se fez necessário a realização de pesquisas e estudos a respeito do processo legislativo brasileiro e todo o âmbito da Câmara dos Deputados.

Compreensão dos dados

A compreensão dos dados abrangeu as etapas de coleta, descrição, exploração e verificação da qualidade dos dados. Foi adotado a legislatura 56 correspondendo ao mandato de 2019 a 2022, com: Conjunto de deputados - 6 atributos e 626 registros; Proposições por ano de apresentação - 11 atributos e 134.337 registros; Proposições por temática - 3 atributos e 59.300 registros e proposição por ano de apresentação - 6 atributos e 303.302 registros. Todos os atributos são descritos na Tab.1.

Análise exploratória

O termo análise exploratória foi criado pelo estatístico John Tukey e consiste na descoberta de fatos e insights, fazendo uso de um determinado procedimento, em que os principais objetivos são visualizar e explorar os dados sob diversos ângulos diferentes. Sendo assim, a forma como a investigação pode ser realizada sob diferentes óticas faz surgir também diferentes conjecturas, de modo que o analista encontre indícios interessantes nos dados coletados [9].

Para o presente trabalho foram selecionadas hipóteses relacionadas às proposições por ano de apresentação, incluindo a classificação temática e os autores.

Hipótese 1: Proposições do tema direitos humanos e minorias são as mais frequentes;

Objetivo: Avaliar a prioridade dada aos temas de direitos humanos e minorias no contexto legislativo, verificando se esses temas são os mais recorrentes nas proposições apresentadas.

Hipótese 2: O estado de São Paulo é o maior em número de proposições apresentadas;

Objetivo: Investigar se São Paulo, como um dos estados mais populosos e economicamente influentes do Brasil, lidera em termos de número de proposições apresentadas à Câmara dos Deputados.

Hipótese 3: Proposições relacionadas à saúde tiveram um aumento significativo durante a pandemia em 2020 e 2021;

Objetivo: Analisar o impacto da pandemia de COVID-19 nas atividades legislativas, verificando se houve um aumento nas proposições relacionadas à saúde durante os anos críticos da pandemia.

Hipótese 4: Proposições apresentadas por deputados de partidos de oposição são mais propensas a serem rejeitadas do que proposições apresentadas por deputados de partidos da base governista;

Objetivo: Examinar a influência da filiação partidária na aprovação de proposições, explorando se existe um viés político que favorece proposições de partidos da base governista em detrimento das de oposição.

Hipótese 5: O partido com maior número de proposições apresentadas foi o PL;

Objetivo: Identificar qual partido político é mais ativo na apresentação de proposições, especificamente verificando se o partido PL é o mais prolífico.

Hipótese 6: Ano eleitoral detém a maior quantidade de proposições apresentadas;

Objetivo: Explorar se há um aumento na atividade legislativa durante anos eleitorais, possivelmente devido a estratégias de reeleição ou maior engajamento dos deputados.

Hipótese 7: Proposições do tipo PL (projetos de Lei) estão entre as 10 mais apresentadas à Câmara dos Deputados;

Objetivo: Verificar a popularidade e frequência dos projetos de lei em comparação com outros tipos de proposições, confirmando se eles estão entre os mais comuns.

Hipótese 8: A maioria das proposições apresentadas são desenvolvidas por Deputados.

Objetivo: Avaliar o papel dos deputados como principais autores de proposições, em comparação com outros atores como comissões ou órgãos legislativos.

Essas hipóteses, em conjunto, buscam fornecer uma visão mais abrangente sobre o comportamento legislativo, a dinâmica política e as prioridades temáticas na Câmara dos Deputados, contribuindo para uma compreensão mais profunda do seu funcionamento.

Assim sendo, para realização da análise exploratória, optou-se por utilizar a linguagem Python e as seguintes bibliotecas: Requests, Json, Pandas, Matplotlib, Seaborn, DateTime, Numpy, Scipy e WordCloud. Além disso, foi utilizada a API disponibilizada pelo portal, tendo em vista a facilidade oferecida quanto à seleção de legislaturas. Dessa forma, os dados foram transformados em data frames com a biblioteca Pandas.

Preparação dos dados

Para a aplicação desta etapa, utilizou-se a linguagem de programação Python, juntamente com as bibliotecas Pandas e Scikit-Learn. Portanto, a primeira ação quanto a preparação de dados, consistiu na escolha do conjunto de proposições para a modelagem.

Consoante ao objetivo proposto pela etapa seguinte, referente à modelagem de dados, buscou-se filtrar as colunas com baixa variância e alta correlação, resultando no conjunto de dados apresentados na Tab.1.

Tabela. 1: Elementos do conjunto de proposições. Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)

Coluna	Descrição
id	Identificador único de cada proposição
siglaTipo	sigla que referencia ao tipo da proposição
numero	número da proposição que a identifica na tabela de classificação temática
ano	ano em que a proposição foi apresentada
codTipo	código que referencia ao tipo da proposição
descricaoTipo	Descrição do tipo da proposição
ementa	Representa o resumo da proposição
ementaDetalhada	Detalha a ementa da proposição
keywords	Palavras chaves do que se trata a proposição
dataApresentacao	Data contendo dia, mês e ano em que a proposição foi apresentada à Câmara dos Deputados
ultimoStatusDescricaoTramitação	Representa o status em que a proposição se encontra, podendo ser aprovada, rejeitada ou arquivada
numero	Número que identifica a proposição
ano	Ano que a proposição foi apresentada
tema	Tema em que a proposição se enquadra
idProposição	Identificador da proposição
idDeputado	Identificador único de deputado autor da proposição
tipoAutor	Tipo do autor
nomeAutor	Nome do parlamentar
siglaPartidoAutor	Sigla do Partido a qual o parlamentar é afiliado
siglaUfAutor	Unidade Federativa a qual o deputado foi eleito

Modelagem dos dados

Para a etapa de modelagem, a biblioteca Scikit-Learn para linguagem de programação Python foi selecionada, pois conta com inúmeras ferramentas para projetos de aprendizado de máquina. Com isso, a classe K-means do módulo sklearn.cluster foi importada. Dessa classe, criou-se um objeto definido com os seguintes parâmetros:

- n_clusters com intervalo [2, 21] - quantidade de clusters em que os dados do conjunto estudado serão agrupados;
- init = ‘k-means++’ - método de inicialização dos centroides de cada cluster. A inicialização do tipo ‘k-means++’ seleciona os centroides iniciais a partir da utilização de amostragem que se baseia em uma distribuição de probabilidade empírica da contribuição de cada registro para a inércia geral;
- n_init = 10 - número de vezes que o k-means será executado com diferentes centroides iniciais;
- max_iter = 300 - quantidade máxima de iterações do k-means em uma única execução, caso a convergência para um agrupamento estável não seja alcançada;
- random_state = 0 - ao passar um valor inteiro garante que os resultados sejam iguais a cada chamada do K-Means.

Ao final da clusterização, uma nova coluna foi acrescida ao dataset, contendo os labels dos clusters, o que permitiu a identificação de qual grupo cada registro se encontrava.

Avaliação

A partir dos labels resultantes de cada clusterizacão e utilizando a biblioteca Scikit-Learn, calculou-se o

coeficiente de silhueta médio do agrupamento. Assim, para cada clusterização variando de 2 a 21, armazenou-se o coeficiente de silhueta médio das modelagens, sendo possível definir o melhor número de clusters para o conjunto de dados sendo igual a 3.

Assim, com a modelagem agrupando em 3 clusters distintos, foram realizadas as análises das características dos elementos de cada cluster de forma a validar as hipóteses levantadas anteriormente. Os resultados dessa etapa são expostos na seção de resultados e discussões.

Implantação

A última etapa do CRISP-DM visa apresentar os conhecimentos obtidos e será executada a tarefa de produção do relatório final, através de técnicas de visualização de dados, utilizando a linguagem Python. Nesse sentido, essa etapa busca sintetizar os resultados, possibilitando comparar com o que era esperado, aprender e expandir o conhecimento adquirido. Portanto, compreende-se como a etapa de implantação, o resultado final da escrita deste trabalho.

IV. RESULTADOS E DISCUSSÃO

As análises referem-se ao conjunto de proposições por ano de apresentação, incluindo a classificação temática e os autores, buscando-se validar ou invalidar as hipóteses definidas anteriormente, sendo detalhadas a seguir:

Hipótese 01: Proposições do tema direitos humanos e minorias são as mais frequentes. Essa hipótese é refutada, uma vez que o tema administração pública é o mais frequente, com 935.326 proposições apresentadas à Câmara dos Deputados, seguido por processo legislativo e atuação parlamentar (745.883), finanças públicas e orçamento (722.077), saúde (446.205), direitos humanos e minorias (369.432), entre outros. Assim, podemos concluir que na legislatura 56 (correspondendo ao mandato de 2019 a 2022) os temas de direitos humanos e minorias no contexto legislativo não foram os mais recorrentes nas proposições apresentadas.
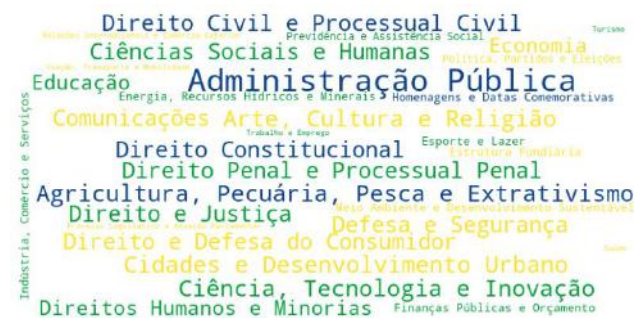
Fig. 1: Temas de proposições apresentados com maior frequência - Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)

Para a validação dessa hipótese, buscou-se desenvolver uma nuvem de palavras conforme a frequência de temas apresentados. Para isso, utilizou-se a biblioteca WordCloud, sendo o resultado demonstrado na Fig.1, em que as palavras diminuem conforme o número de proposições também diminui.

Hipótese 02: O estado de São Paulo é o maior em número de proposições apresentadas. Utilizando o método groupby para agrupar os dados de acordo com o id da proposição e a sigla federativa do autor, obteve-se como resultado um novo dataframe com os estados e a quantidade de proposições apresentadas. Dessa forma, São Paulo foi o estado líder em proposições apresentadas, somando 426.460 apresentações à câmara, sendo essa hipótese, portanto, verdadeira. Outrossim, a ideia de tal implementação foi a de que o maior estado teria a maior quantidade de deputados, o que foi comprovado pela análise exploratória do conjunto de deputados e, portanto, teria o maior número de proposições apresentadas. Para melhor visualização, o gráfico de barras, Fig.2, demonstra o resultado obtido.
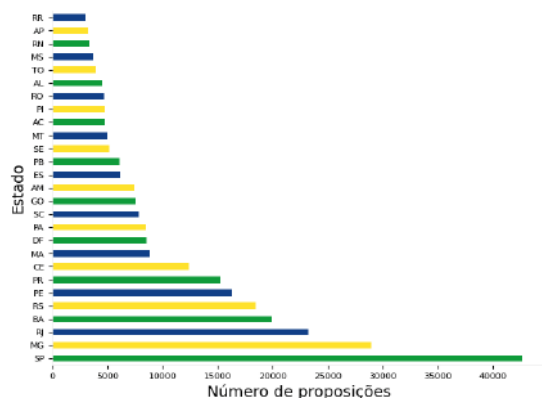
Fig. 2: Estados com maior frequência de proposições apresentadas à Câmara dos Deputados - Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)

Hipótese 03: Proposições relacionadas à saúde tiveram um aumento significativo durante a pandemia em 2020 e 2021. Apesar do contexto pandêmico, a análise mostra que maiores quantidades de proposições relacionadas à saúde foram apresentadas no ano de 2022, quando a pandemia já estava reduzida. No entanto, podemos notar um aumento de aproximadamente 42% das proposições pós-pandemia. Como o objetivo central dessa hipótese era analisar o impacto da pandemia de COVID-19 nas atividades legislativas, os dados demonstram que a maioria das proposições são relacionadas após o momento mais crítico da pandemia. A Fig.3 auxilia no entendimento dessas dimensões.
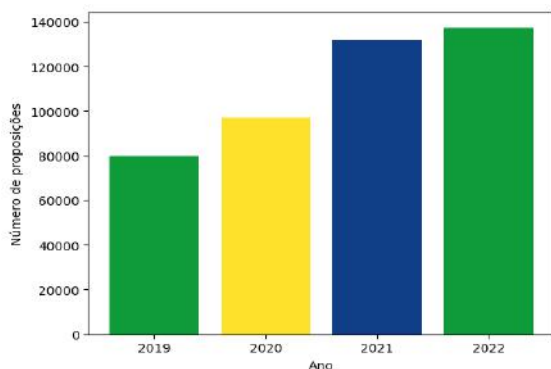
Fig. 3: Proposições relacionadas à saúde por ano de apresentação - Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)

Hipótese 04: Proposições apresentadas por deputados de partidos de oposição são mais propensas a serem rejeitadas do que proposições apresentadas por partidos da base governista. Para validação dessa hipótese, foi desenvolvido um gráfico, Fig.4, com a proporção das proposições aprovadas e rejeitadas para cada um dos partidos, comprovando o fato de que as proposições da base são mais aprovadas do que proposições de oposição. Tais dados corroboram com a hipótese da influência da afiliação partidária na aprovação de proposições, podendo indicar uma maior facilidade para aprovação de proposições de partidos da base governista em detrimento das de oposição.
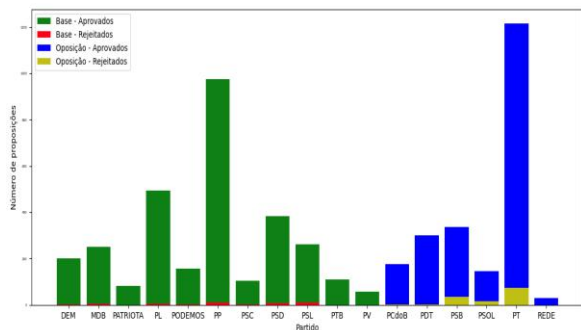
Fig. 4: Porção de proposições aprovadas e rejeitadas por partido - Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)

Hipótese 05: O partido com maior número de proposições apresentadas foi o PL. Sabendo que a hipótese anterior foi verdadeira, em que partidos da base governista obtiveram maior quantidade de proposições aprovadas, pode ser possível também validar a hipótese de que o partido da base vigente também teria o maior número de proposições apresentadas à câmara. Todavia, o PT foi o que mais deteve autores de proposições apresentadas com um total de 64.116, seguido por PSB com 22.059, PDT com 18.682, PL com 14.644, dentre outros partidos de acordo com a Fig.5.
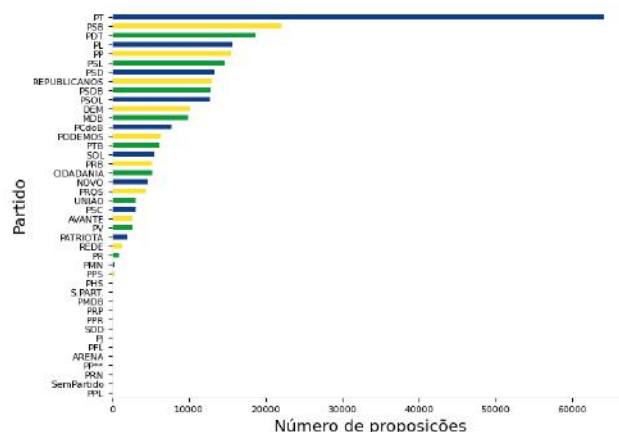
Fig. 5: Proposições apresentadas por partido - Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)

Hipótese 06: Ano eleitoral detém a maior quantidade de proposições apresentadas. Para implantação dessa hipótese, considerou-se a tentativa de reeleição de muitos deputados e uma das estratégias poderia ser maior participação nas atividades legislativas, sendo a apresentação de proposições uma delas. Contudo, o resultado dessa afirmação é falso, uma vez que o ano eleitoral (2022) ficou em terceiro lugar com 26.774 proposições apresentadas, antecedido por 2019 com 41.200, 2021 com 39.888 e sucedido por 2020 com 26.475. Assim, os resultados são demonstrados no gráfico da Fig.6.
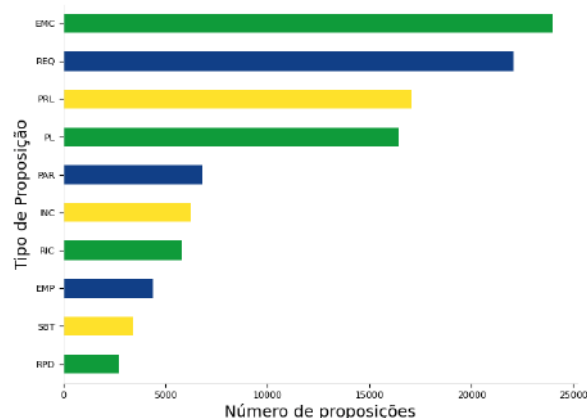
Fig. 6: Proposições apresentadas por ano - Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)

Hipótese 07: Proposições do tipo PL (Projetos de Lei) estão entre as 10 mais apresentadas à Câmara dos Deputados. Essa hipótese é verdadeira, uma vez que projetos de lei contaram com 16.462 proposições apresentadas, sendo antecedido apenas por EMC (Emenda Constitucional) com 23.975, REQ (Requerimento de Sessão Solene) com 22.042 e PRL (Parecer do Relator) com 17.079, Fig.7. Tais dados comprovaram a popularidade e frequência dos projetos de lei em comparação com outros

tipos de proposições, confirmando se eles estão entre os mais comuns.
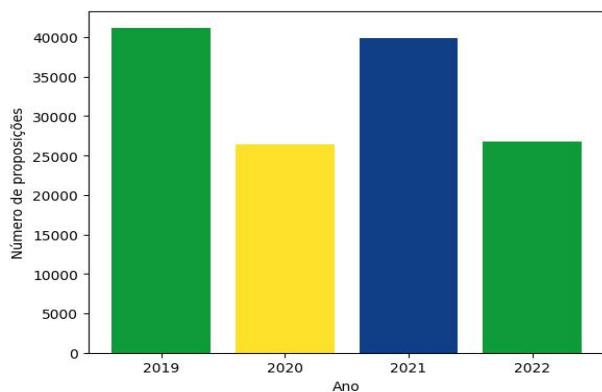
Fig. 7: Porção de proposições por tipo - Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)
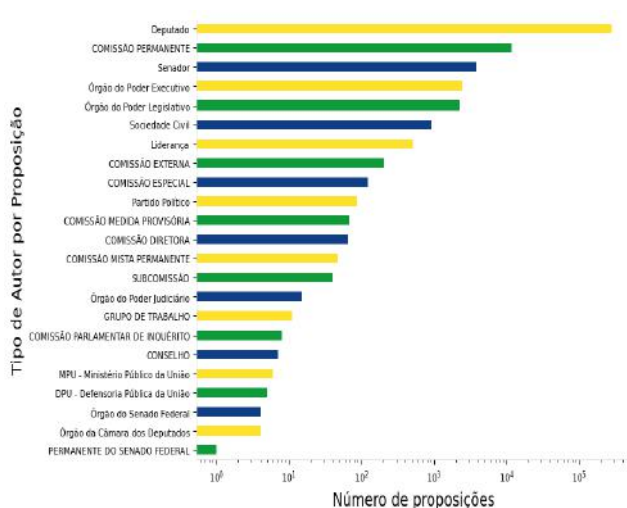
Fig. 8: Porção de proposições por tipo de autor - Legislatura 56. Fonte: Elaborado pelo autor(a) (2025)

Hipótese 08: A maioria das proposições apresentadas são desenvolvidas por deputados. Para validação dessa hipótese, foi necessário entender o conjunto sobre autores, uma vez que não são apenas deputados, sendo possível a autoria por meio de órgãos legislativos, comissão externa, comissão permanente, senador, sociedade civil, dentre outros. Assim, por se tratar do ambiente da câmara de deputados, assume-se que deputados seriam a maior categoria no que diz respeito à apresentação de proposições, sendo essa uma verdade. Deputados ocupam a maior posição de autorias, contando com 279.829 proposições apresentadas, seguidos por comissão permanente com 11.772, Senador 3.809, dentre outros. A Fig.8 auxilia na visualização dessa hipótese.

Os resultados encontrados revelam as proporções dos dados referentes às proposições em relação aos seus principais atributos, auxiliando no entendimento e acompanhamento da ocorrência de cada uma das atividades legislativas em âmbitos como federação, partidos, ano, tema, dentre outros. Nesse sentido, o aumento constante de dados em todo o mundo oferece diversas possibilidades de utilização, o que também se aplica ao contexto da Câmara dos Deputados.

Desse modo, o cenário levantado nessa pesquisa aborda a dificuldade encontrada em obter conhecimentos acerca da conjuntura legislativa e foi nesse viés, que os autores buscaram demonstrar, de forma mais acessível e compreensível, as relações entre os dados legislativos, a partir do levantamento de hipóteses, explorando cada um dos registros coletados. Além disso, é perceptível que esse cenário é resultado da carência de informações por parte da população, ainda que com a disponibilidade das informações por meio dos portais de transparência, configurando a necessidade de otimizar a forma como o conhecimento é adquirido, uma vez que é influenciado pela forma como os dados estão sendo representados.

V. CONCLUSÃO

Com base nos resultados apresentados, conclui-se que o objetivo central deste trabalho - ampliar a compreensão das atividades legislativas, revelando padrões e relações que possam informar melhor o público e os tomadores de decisão - foi alcançado.

Através da aplicação do algoritmo k-means e o modelo Crisp-DM, foi possível analisar a veracidade, ou não, de um conjunto de hipóteses relacionadas às proposições apresentadas na Câmara dos Deputados.

Este estudo trouxe contribuições nas áreas tecnológica e social. No âmbito tecnológico, apresenta-se uma metodologia de tratamento de dados, baseada na similaridade dos mesmos, disponíveis no portal de dados abertos da Câmara dos Deputados. Para o âmbito social, a principal contribuição constitui-se em uma maior transparência e entendimento, por parte da população, a respeito das proposições dos parlamentares eleitos, visando confirmar se de fato o seu trabalho está sendo bem desempenhado.

Outrossim, para as pretensões acadêmicas e profissionais, este trabalho representa um passo significativo. Profissionalmente, este estudo proporcionou uma base sólida de análise de dados e compreensão das dinâmicas do poder legislativo no Brasil, oferecendo novas perspectivas e metodologias que podem ser exploradas em pesquisas futuras.

Por fim, o presente estudo consistiu em demonstrar uma das mais variadas possibilidades da análise dos dados legislativos e por isso, é válido direcioná-lo como pontapé

para trabalhos futuros, que possam colaborar ainda mais para a obtenção de resultados voltados à análise de dados da estrutura política brasileira.

REFERÊNCIAS

[1] Rowley, J. The wisdom hierarchy: Representations of the dikw hierarchy, (2007). *Journal of Information Science*, v.33, 163-180. ISSN 01655515

[2] Neto, B. P. et al. Publicidade e Transparência das Contas Públicas: obrigatoriedade e abrangência desses princípios na administração pública brasileira, 2007, from http://www.redalyc.org.articulo.oa?id=197014728005

[3] Perna, A. S. (2010). O Lado Invisível da Participação Política: Gestão da Informação dos Mecanismos Digitais de Participação Política nos Parlamentos da América Latina, com uma Análise do Caso da Câmara dos Deputados do Brasil. *Dissertação de Mestrado - Universidade de Brasília*

[4] Meijer, A. Understanding the complex dynamics of transparency. Public Administration Review, (2013). v. 73, p. 429–439, 5 2013. ISSN 00333352

[5] Lyrio, M., Lunkes, R. & Castello-Taliani, E. (2019). Transparência Governamental na Internet: Uma Análise Comparativa no Âmbito do Poder Executivo Brasileiro e Espanhol 1

[6] Warzocha, G. & Cruvinel, F. (2016). Câmara dos Deputados Centro de Formação, Treinamento e Aperfeiçoamento. Programa de Pós-graduação - Mestrado Profissional em Poder Legislativo

[7] Cardoso, P. H. (2019). Ciência de dados aplicada a dados governamentais abertos sob a ótica da Ciência da Informação

[8] Shearer, C. The crisp-dm model: The new blueprint for data mining, (2000). Journal of Data Warehousing, v. 5, n. 4, p. 13–22, from <https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf>

[9] Tukey, J. W. Exploratory Data Analysis. [S.l.]: Addison-Wesley Publishing Company, 1977, 7-710 p.